
Automated Quantification of ^{18}F -Flutemetamol PET Activity for Categorizing Scans as Negative or Positive for Brain Amyloid: Concordance with Visual Image Reads

Lennart Thurfjell^{1,2}, Johan Lilja¹, Roger Lundqvist¹, Chris Buckley³, Adrian Smith³, Rik Vandenberghe^{4,5}, and Paul Sherwin⁶

¹GE Healthcare, Uppsala, Sweden; ²Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden; ³GE Healthcare, London, United Kingdom; ⁴Laboratory for Cognitive Neurology, Catholic University Leuven, Leuven, Belgium; ⁵Neurology Department, University Hospitals Leuven, Leuven, Belgium; and ⁶GE Healthcare, Princeton, New Jersey

Clinical trials of the PET amyloid imaging agent ^{18}F -flutemetamol have used visual assessment to classify PET scans as negative or positive for brain amyloid. However, quantification provides additional information about regional and global tracer uptake and may have utility for image assessment over time and across different centers. Using postmortem brain neuritic plaque density data as a truth standard to derive a standardized uptake value ratio (SUVR) threshold, we assessed a fully automated quantification method comparing visual and quantitative scan categorizations. We also compared the histopathology-derived SUVR threshold with one derived from healthy controls. **Methods:** Data from 345 consenting subjects enrolled in 8 prior clinical trials of ^{18}F -flutemetamol injection were used. We grouped subjects into 3 cohorts: an autopsy cohort ($n = 68$) comprising terminally ill patients with postmortem confirmation of brain amyloid status; a test cohort ($n = 172$) comprising 33 patients with clinically probable Alzheimer disease, 80 patients with mild cognitive impairment, and 59 healthy volunteers; and a healthy cohort of 105 volunteers, used to define a reference range for SUVR. Visual image categorizations for comparison were from a previous study. A fully automated PET-only quantification method was used to compute regional neocortical SUVRs that were combined into a single composite SUVR. An SUVR threshold for classifying scans as positive or negative was derived by ranking the PET scans from the autopsy cohort based on their composite SUVR and comparing data with the standard of truth based on postmortem brain amyloid status for subjects in the autopsy cohort. The derived threshold was used to categorize the 172 scans in the test cohort as negative or positive, and results were compared with categorization using visual assessment. Different reference and composite region definitions were assessed. Threshold levels were also compared with corresponding thresholds derived from the healthy group. **Results:** Automated quantification (using pons as the reference region) demonstrated 91% sensitivity and 88% specificity and gave 3 false-positive and 4 false-negative scans. All 3 false-positive cases were either borderline-normal by standard of truth or had moderate to heavy cortical diffuse plaque burden. In the

test cohort, the concordance between quantitative and visual read categorization ranged from 97.1% to 99.4% depending on the selection of reference and composite regions. The threshold derived from the healthy group was close to the histopathology-derived threshold. **Conclusion:** Categorization of ^{18}F -flutemetamol amyloid imaging data using an automated PET-only quantification method showed good agreement with histopathologic classification of neuritic plaque density and a strong concordance with visual read results.

Key Words: quantification; ^{18}F -flutemetamol; amyloid imaging

J Nucl Med 2014; 55:1623–1628
DOI: 10.2967/jnumed.114.142109

Amyloid PET radiotracers have been developed for in vivo assessment of amyloid β deposition in the brain and may prove useful both clinically and in Alzheimer disease research. The first widely used research amyloid PET tracer was the ^{11}C -labeled investigational tracer Pittsburgh compound B (*1*). However, the 20-min half-life of ^{11}C restricts the use of Pittsburgh compound B to centers with an on-site cyclotron. For wider access and commercialization, a tracer with a longer half-life is needed, and several tracers labeled with ^{18}F (half-life, 110 min) are being developed. At the time of writing, ^{18}F -florbetapir (*2,3*) and ^{18}F -florbetaben (*4,5*) are approved for clinical use by the Food and Drug Administration and the European Medicines Agency, and ^{18}F -flutemetamol (*6,7*) is approved by the Food and Drug Administration and is under evaluation by the European Medicines Agency. Another compound in late-stage development is the investigational amyloid imaging agent NAV4694 (*8*).

The currently approved amyloid imaging PET tracers have been approved for visual assessment of the PET image in which a trained reader categorizes scan results as negative (normal uptake) or positive (increased gray-matter uptake). However, quantification may be important for several reasons. Recent work indicates that the amount of amyloid has a prognostic value (*9*) and that any assessment of longitudinal change requires quantification. Quantification also facilitates comparison of results across centers. Quantitative methods often rely on computation of standardized uptake value ratios (SUVRs) between target regions and a reference region in a late-sum

Received Apr. 30, 2014; revision accepted Jul. 22, 2014.
For correspondence or reprints contact: Lennart Thurfjell, MDx PET Biomarkers and Software, GE Healthcare/Uppsala Imanet, P.O. Box 967, SE-751 09 Uppsala, Sweden.
E-mail: lennart.thurfjell@ge.com
Published online Aug. 21, 2014.
COPYRIGHT © 2014 by the Society of Nuclear Medicine and Molecular Imaging, Inc.

scan (10). To use quantitative data for categorization of scans as positive or negative, an SUVR threshold is needed. Previously proposed methods for threshold derivation include methods based on receiver-operating-curve (ROC) analysis (11,12) and a method based on the statistical distance between Alzheimer disease and healthy volunteer scans (7). If standard-of-truth data are available, an SUVR threshold that best separates subjects with and without brain amyloid can be found, as proposed by Clark et al. (3). Another approach, frequently used for ^{18}F -FDG, applies data from healthy volunteers to generate a reference SUVR database from which a threshold can be derived.

In this paper, we quantified ^{18}F -flutemetamol scans using a fully automated method and used postmortem data to derive an SUVR threshold. Using the threshold, we categorized scan results as positive or negative and compared these categorizations with those made by blinded readers in a previous clinical trial. We also assessed the usefulness of an ^{18}F -flutemetamol reference database, with a threshold based on mean and SD information derived from amyloid-negative ^{18}F -flutemetamol scans being used to categorize scans as positive or negative. We compared these results with those resulting from the postmortem standard-of-truth-based threshold.

MATERIALS AND METHODS

Subjects and Imaging

Data from 345 subjects imaged in 8 previous clinical studies of ^{18}F -flutemetamol were used. All participants gave written informed consent that was approved by the Internal Review Board of each participating institution. We grouped subjects into 3 cohorts: an autopsy cohort ($n = 68$) comprising terminally ill patients with subsequent postmortem confirmation of brain amyloid status; a test cohort ($n = 172$) comprising 33 patients with clinically probable Alzheimer disease, 80 patients with mild cognitive impairment, and 59 healthy volunteers; and a healthy cohort comprising 105 healthy volunteers from 3 studies (7,13,14).

Data in the autopsy cohort were from a phase III, multicenter PET study of ^{18}F -flutemetamol injection for detecting brain amyloid β (ClinicalTrials.gov identifier NCT01165554). Data in the test cohort were from a phase III study in which the effectiveness of an electronic training program for interpretation of ^{18}F -flutemetamol images was assessed (ClinicalTrials.gov identifier NCT01672827).

Except in the autopsy cohort, ^{18}F -flutemetamol PET images had been previously acquired as six 5-min frames starting 85–90 min after injection of approximately 185 MBq of ^{18}F -flutemetamol; for this study, the first 4 frames were summed to yield a 20-min static scan. In the autopsy trial, the dose was between 185 and 370 MBq at investigator discretion, to allow use of higher doses to shorten imaging time for subjects who could not tolerate a 20- to 30-min imaging time because of their terminal condition. End-of-life subjects who were prone to agitation could also undergo sedation for the PET scan

according to local clinical practice. Most end-of-life subjects underwent dynamic scanning in 2-min frames starting 90 min after injection. The first frames (typically 5) were summed to generate an image comprising the first 10 min of the scan.

Histopathology Analysis

The brains of subjects in the autopsy cohort were analyzed as part of a phase III study (ClinicalTrials.gov identifier NCT01165554). Histopathology data were obtained from subjects who died during the study and underwent brain autopsy. Postmortem histopathologic examination was performed at a central pathology laboratory. Two tissue blocks were taken from each of the following regions: precuneus, midfrontal cortex, superior temporal cortex, middle temporal cortex, inferior parietal cortex, anterior cingulate gyrus, posterior cingulate gyrus, and primary visual cortex. Three slides were prepared from tissue slices taken from each block, for a total of 6 slides per region. Each slide was stained using the Bielschowsky silver stain (15). Each of 5 fields of view per slide was assessed for neuritic plaque density, which was classified as none (0 plaques), sparse (1–5 plaques), moderate (6–19 plaques), or frequent (≥ 20 plaques), following Vemuri's modification of the Consortium to Establish a Registry for Alzheimer Disease criteria (16,17). Each field-of-view assessment was scored as 0 (for none), 1 (for sparse), 2 (for moderate), or 3 (for frequent), and the 5 field-of-view scores were averaged for each of the 6 tissue slides prepared from a region. The 6 mean field-of-view scores were then averaged to give an overall mean Bielschowsky score for each of the 8 cortical regions. The midpoint (1.5) of the neuritic plaque density scoring scale was used as the threshold for categorizing overall mean Bielschowsky score as normal or abnormal. If the mean Bielschowsky score was greater than 1.5, the region was classified as abnormal, and the subject's brain was classified as abnormal. The brain was classified as normal only if all regions were classified as normal. The brain classification was used as the standard of truth for determining sensitivity and specificity.

Blinded Image Evaluation

Masked image evaluations were conducted previously in a phase III study in which the effectiveness of an electronic training program for interpretation of ^{18}F -flutemetamol images was assessed (ClinicalTrials.gov identifier NCT01672827). In this study, each subject's PET image was interpreted separately by 5 readers, each of whom was blinded to all subject clinical information and the neuropathology classification. Readers had been trained using a DVD-based program. Images were reviewed in color, typically using a Sokolov or rainbow color scale. The readers were trained to assess the following regions for ^{18}F -flutemetamol activity: frontal, posterior cingulate/precuneus, lateral temporal, inferior parietal, and striatal. Positive images were defined as having at least one cortical region with reduction or loss of the normally distinct gray matter–white matter contrast, that is, one or more regions with increased critical gray matter ($>50\%$ – 60% intensity) or reduced white matter contrast (the white matter sulcal pattern being less distinct). The results of the blinded image evaluations were available for all scans in the autopsy and test cohorts. For this study, we used the majority visual read results, that is, the scan classification on which at least 3 readers agreed.

Quantification

^{18}F -flutemetamol activity was quantified using a previously described fully automated PET-only method that uses an adaptive template for handling different uptake patterns in negative and positive flutemetamol images (18). The adaptive template method was used

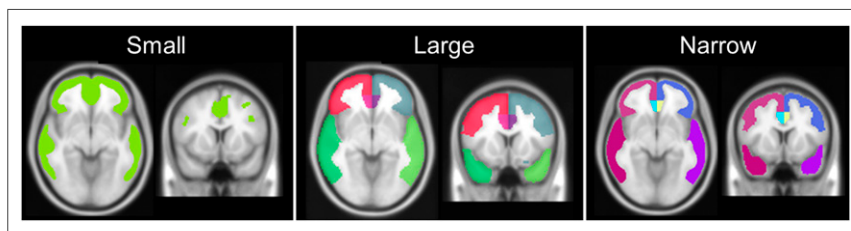


FIGURE 1. The 3 composite regions outlined on the Montreal Neurological Institute average T1 template.

TABLE 1
PET Quantitative Categorization vs. Standard of Truth for Different Target and Reference Regions

Reference region	Composite region	SUVr Threshold	Sensitivity	Specificity	AUC
Pons	Small	0.62	91%	88%	0.93
	Narrow	0.58	91%	84%	0.93
	Large	0.56	91%	84%	0.92
Cerebellar gray matter	Small	1.67	86%	88%	0.90
	Narrow	1.57	86%	88%	0.89
	Large	1.63	86%	88%	0.87
Whole cerebellum	Small	1.30	91%	84%	0.92
	Narrow	1.27	88%	84%	0.92
	Large	1.23	88%	84%	0.90

to spatially normalize the flutemetamol scans to Montreal Neurological Institute template space. Retention ratios were computed for a cortical composite region sampling areas from the frontal, parietal, lateral temporal, and anterior and posterior cingulate cortices. We evaluated 3 types of cortical composite region: small, large, and narrow. A small composite region was defined as the areas where the difference between healthy volunteer and Alzheimer patient is maximal. A large composite region was defined as a combination of automated anatomic labeling-type regions (19) outlined on the ICBM-152 template masked with a gray matter probability mask. A narrow composite region was defined as being the same as the large region but with narrowing to increase the distance both from the gray matter–white matter border and from the brain surface to the region (Fig. 1). The reference regions were pons, cerebellar gray matter, and whole cerebellum, and retention ratios were computed for all combinations of composite and reference regions.

Experiments

The 68 composite SUVr for the autopsy cohort were ranked and compared with the standard-of-truth data. An ROC analysis was performed, and the area under the curve (AUC) was used as a measure of the method’s ability to correctly categorize scans. Furthermore, we identified the SUVr threshold that best categorized scans, and sensitivity and specificity were calculated. The threshold was then used to categorize the scans in the test cohort, and the results were compared with the previous visual read results. Computation of AUC, derivation of the SUVr threshold, and subsequent categorization of scans and comparison with visual read results were done for all combinations of reference and cortical composite regions.

As an independent way of computing the threshold, we used the healthy cohort to derive an SUVr threshold. When individual scans are compared with a reference database, z scores are typically computed as follows:

$$z = ([\text{individual value}] - [\text{normal mean}]) / (\text{normal SD}). \quad \text{Eq. 1}$$

We investigated the use of a threshold corresponding to $z = 2.0$ and $z = 2.5$, where z defines the number of SDs above the normal mean. Thresholds derived from the reference database were compared with the corresponding thresholds derived from the autopsy cohort. This comparison was done for all reference regions but only for the narrow cortical composite region.

Statistical Analysis

Statistical analysis was performed using R (20), and ROC analysis was performed using the pROC plug-in (21).

RESULTS

Comparison with Histopathology

ROC analysis showed that use of the pons as the reference region separated scans from standard-of-truth–normal and standard-of-truth–abnormal subjects, with an AUC ranging from 0.92 to 0.93 depending on the composite region used. When whole cerebellum and cerebellar gray matter were used as reference regions, the corresponding ranges were 0.90–0.92 and 0.87–0.90, respectively (Table 1). Pairwise significance testing using the method reported by DeLong et al. (22) showed that the only significant differences ($P < 0.05$) between the AUC when comparing reference regions were whole cerebellum versus cerebellar gray matter for the large and narrow target regions. The corresponding significance testing when comparing AUC within the same reference region but between target regions showed significant differences ($P < 0.05$) between small and large regions only when cerebellar gray matter was used as the reference region.

With pons as the reference region and using the small target region, an SUVr of 0.62 gave the best separation of scans from standard-of-truth–normal and standard-of-truth–abnormal subjects (Fig. 2). This threshold gave 3 false-positive scans and 4 false-negative scans, yielding a sensitivity of 91% and a specificity of 88%. Comparison with visual read results showed that all false-positive scans were positive also by visual read and 3 of 4 false-negative scans were negative by visual read (Fig. 3). The narrow and large target regions gave similar results but with slightly lower

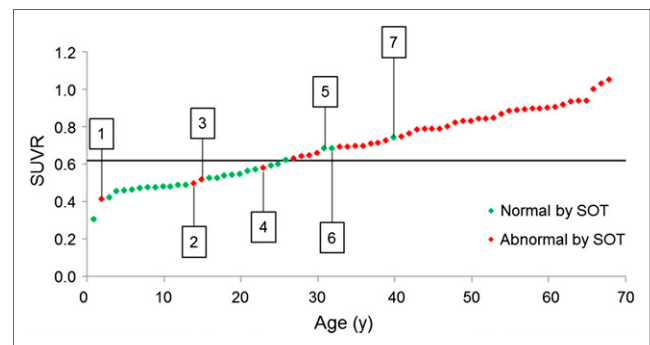


FIGURE 2. Plot showing SUVr data (pons as reference region) from 68 subjects in autopsy cohort in ascending order from left to right. Line corresponds to threshold that best separates standard-of-truth–normal from standard-of-truth–abnormal subjects (0.62). False-negative and false-positive scans are indicated by boxed numbers.

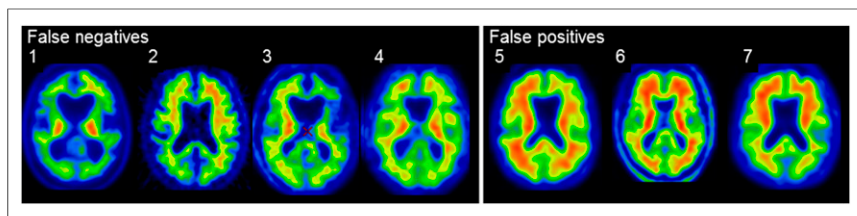


FIGURE 3. The 3 false-positive and 4 false-negative scan obtained using Pons-based SUVR threshold of 0.62. The 3 false-positive scans were abnormal and 3 of 4 false-negative scans were normal by visual read.

SUVr thresholds (0.58 and 0.56, respectively) for the pons reference region. The results for all combinations of target and reference regions are summarized in Table 1.

Categorization of Scans

Dichotomous categorization of scans in the test cohort as negative or positive using the thresholds derived from the autopsy cohort gave concordance with visual read results ranging from 97.1% to 99.4% depending on the combination of reference and cortical composite regions. Figure 4 shows a plot illustrating how the autopsy-derived threshold for the pons/narrow region combination accurately separated 171 of 172 scans in the test cohort. Results for all region combinations summarized in Table 2 show that the concordance for quantitative and visual categorization was similar across the different composite regions. Comparison of the different reference region methods showed that pons and whole cerebellum-normalized data gave similar results (98.8%–99.4%), whereas data normalized to cerebellar gray matter had slightly lower concordance with visual read results (97.1%–98.8%).

Reference Database Cohort

SUVr thresholds corresponding to a z score of 2.0 and 2.5 in the healthy cohort and with the narrow composite region are summarized in Table 3. For data normalized to pons, the threshold for $z = 2.0$ was similar to the autopsy-derived threshold (0.59 vs. 0.58) but was slightly lower for data normalized to cerebellar gray matter (1.46 vs. 1.57) and to whole cerebellum (1.21 vs. 1.27). SUVr thresholds corresponding to $z = 2.5$ were slightly higher, but the difference in thresholds $z = 2.0$ versus $z = 2.5$ had only a marginal impact on concordance with visual read results (Table 3).

DISCUSSION

Using an autopsy cohort for which brain amyloid status was available based on a histopathologic standard of truth, we investigated a quantitative method to separate scans from standard-of-truth-normal and standard-of-truth-abnormal subjects. We performed an ROC analysis and computed the AUC for each combination of target and reference region. We determined SUVr thresholds, which we used for categorization of ^{18}F -flutemetamol scans in a test cohort as negative or positive, and we calculated sensitivity and specificity. We compared results from this quantitative categorization with visual majority read results. Furthermore, we compared the thresholds obtained by comparison with histopathology with thresholds determined on the basis of the distribution in a group of amyloid-negative healthy control subjects.

For the autopsy cohort, group separation when using pons as the reference region gave AUC values in the range of 0.92–0.93.

Categorization of scans using the small composite region gave 3 false-positive and 4 false-negative results. The 3 false-positive were all positive by visual read, and 3 of the 4 false-negative were negative by visual read. Analysis of the autopsy cohort data showed that the 4 false-negative had a neuritic plaque score that was close to the standard-of-truth threshold, suggesting these were borderline cases. Each also had some degree of cortical atrophy. Two false-positive results occurred in patients with a clinical history of dementia and neuropathologic

evidence consistent with dementia with Lewy bodies who had diffuse plaques, as well as sparse or moderate neuritic plaques; these were below the 1.5 mean Bielschowsky score threshold for classifying the brains as abnormal for amyloid.

Use of the whole cerebellum as the reference region gave AUC values in the range of 0.90–0.92, and use of cerebellar gray matter as the reference region gave AUC values in the range of 0.87–0.90. However, the only statistically significant difference was when whole cerebellum was compared with cerebellar gray matter for the large and the narrow target regions. This result may be due to the fact that the autopsy cohort in this study included subjects with severely atrophied brains, which will add more variability to data normalized to the cerebellar gray matter reference region than to data normalized to the other reference regions. The whole-cerebellum reference region was defined as the cerebellar gray matter region with an extension to include also white matter. It seems as if inclusion of white matter partly compensates for the potential deficiencies with a cerebellar gray matter reference region. The choice of reference region will also affect the magnitude of SUVr levels. The pons is a hot region in ^{18}F -flutemetamol scans, and consequently, normalizing against this region will give lower SUVrs than when cerebellar gray matter is used—a region with low ^{18}F -flutemetamol uptake. Whole cerebellum includes both cold and

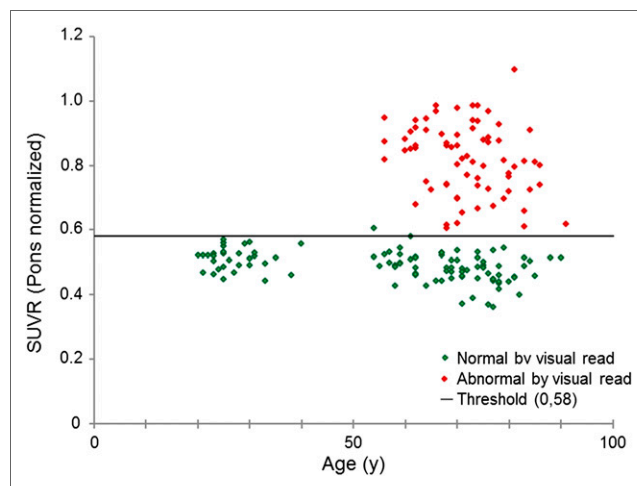


FIGURE 4. SUVR values for all 172 subjects in test cohort plotted against age and color coded on the basis of majority visual read assessment. Pons/narrow SUVR threshold (0.58) derived from autopsy cohort gave concordance between quantitative and visual categorization in 171 of 172 scans.

TABLE 2
Concordance between Quantitative and Visual
Categorization

Reference region	Composite region	No. of concordant scans	% concordance
Pons	Small	171	99.4%
	Narrow	171	99.4%
	Large	170	98.8%
Cerebellar gray matter	Small	170	98.8%
	Narrow	169	98.3%
	Large	167	97.1%
Whole cerebellum	Small	171	99.4%
	Narrow	170	98.8%
	Large	170	98.8%

hot areas, giving SUVRs that are in between SUVRs normalized to pons and cerebellar gray matter.

There were only small differences in results for the different composite regions, both for the comparison with histopathology and for the comparison with visual read results. Results in categorization of the data were almost identical for the small and narrow composite regions but were only slightly inferior for the large composite region (Tables 1 and 2) despite the fact that the composite regions sample quite different areas of the brain (Fig. 1). However, the magnitude of SUVR levels, and hence also the optimal SUVR threshold, vary not only across different reference regions but also across different target regions. The small composite region samples mainly areas where the highest SUVR is expected to be found in amyloid β -positive subjects, whereas the narrow and large composite regions cover larger parts of the neocortex, the difference being that the narrow composite region samples a thin strip of the cortex whereas the large composite region samples most of the cortex and the latter will hence be more affected by partial-volume effects. Therefore, it is expected that SUVRs, and therefore also the SUVR threshold, should be higher for the small composite region than for the large composite region, with the narrow composite region falling in between; this expectation is in agreement with our results (Table 1).

An interesting finding was that our thresholds for the cerebellar gray matter reference region are close to thresholds obtained

with a different, previously reported method (7), and several publications for ^{11}C -labeled Pittsburgh compound B have established SUVR thresholds that are close to 1.5 (11,12). However, although thresholds are similar across different analysis methods, they still vary and a correlation analysis between different methods is necessary for establishing how a threshold using one particular tracer or analysis method translates to a threshold using another tracer or method. This approach was used by Landau et al. (23), who compared 3 tracers and 2 analysis methods. Using correlation analysis between methods, a previously reported threshold for ^{18}F -florbetapir of 1.11 (3) was first converted to ^{11}C -labeled Pittsburgh compound B units and then to ^{18}F -flutemetamol units (whole cerebellum normalization) yielding an ^{18}F -flutemetamol threshold of 1.20, which is similar to the threshold of 1.23 reported for the large composite region (Table 1). Although conversion of SUVR thresholds across tracers and analysis methods can be done using correlation analysis, this is not an ideal approach for clinical routine use. A way to circumvent this problem is to define thresholds in terms of z scores, that is, number of SDs from the mean in a reference database of amyloid-negative controls. Our results showed that classification using a threshold corresponding to $z = 2.0$ gave essentially the same concordance with visual read results as when thresholds were defined on the basis of the autopsy cohort. Moreover, use of a threshold corresponding to $z = 2.5$ only marginally changed the results (Table 3). This finding supports the use of reference databases for flutemetamol in the same way as has previously been used for ^{18}F -FDG (24).

CONCLUSION

Quantification of ^{18}F -flutemetamol amyloid imaging data using an automated PET-only method and categorization of scans using an SUVR threshold derived from an autopsy cohort gave excellent concordance with majority visual read results. The use of pons as the reference region gave a slightly higher AUC and stronger concordance with visual read than did the use of whole cerebellum and cerebellar gray matter, but differences were small, indicating that all 3 reference regions are suitable for evaluation of ^{18}F -flutemetamol PET scans.

DISCLOSURE

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734. No potential conflict of interest relevant to this article was reported.

TABLE 3
Thresholds Obtained Using Amyloid-Negative Healthy Subjects

Reference region	Reference database values		Threshold $z = 2.0$		Threshold $z = 2.5$	
	Mean	SD	SUVR	Concordance with visual read	SUVR	Concordance with visual read
Pons	0.499	0.044	0.59	171/172 (99.4%)	0.61	170/172 (98.8%)
Cerebellar gray matter	1.24	0.11	1.46	166/172 (96.5%)	1.52	169/172 (98.3%)
Whole cerebellum	1.045	0.08	1.21	171/172 (99.4%)	1.25	170/172 (98.8%)

REFERENCES

1. Klunk WE, Engler H, Nordberg A, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh compound-B. *Ann Neurol*. 2004;55:306–319.
2. Wong DF, Rosenberg PB, Zhou Y, et al. In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand ¹⁸F-AV-45 (florbetapir F 18). *J Nucl Med*. 2010;51:913–920.
3. Clark CM, Schneider JA, Bedell BJ, et al. Use of florbetapir-PET for imaging beta-amyloid pathology. *JAMA*. 2011;305:275–283.
4. Barthel H, Gertz HJ, Dresel S, et al. Cerebral amyloid-beta PET with florbetaben (¹⁸F) in patients with Alzheimer's disease and healthy controls: a multicentre phase 2 diagnostic study. *Lancet Neurol*. 2011;10:424–435.
5. Villemagne VL, Ong K, Mulligan RS, et al. Amyloid imaging with ¹⁸F-florbetaben in Alzheimer disease and other dementias. *J Nucl Med*. 2011;52:1210–1217.
6. Nelissen N, Van Laere K, Thurfjell L, et al. Phase I study of the PIB derivative ¹⁸F-flutemetamol in healthy volunteers and patients with probable Alzheimer's disease. *J Nucl Med*. 2009;50:1251–1259.
7. Vandenberghe R, Van Laere K, Ivanoiu A, et al. ¹⁸F-flutemetamol amyloid imaging in Alzheimer disease and mild cognitive impairment: a phase 2 trial. *Ann Neurol*. 2010;68:319–329.
8. Rowe CC, Pejoska S, Mulligan RS, et al. Head-to-head comparison of ¹¹C-PiB and ¹⁸F-AZD4694 (NAV4694) for beta-amyloid imaging in aging and dementia. *J Nucl Med*. 2013;54:880–886.
9. Rowe CC, Bourgeat P, Ellis KA, et al. Predicting Alzheimer disease with beta-amyloid imaging: results from the Australian imaging, biomarkers, and lifestyle study of ageing. *Ann Neurol*. 2013;74:905–913.
10. Lopresti BJ, Klunk WE, Mathis CA, et al. Simplified quantification of Pittsburgh compound B amyloid imaging PET studies: a comparative analysis. *J Nucl Med*. 2005;46:1959–1972.
11. Jagust WJ, Landau SM, Shaw LM, et al. Relationships between biomarkers in aging and dementia. *Neurology*. 2009;73:1193–1199.
12. Ng SY, Villemagne VL, Masters CL, Rowe CC. Evaluating atypical dementia syndromes using positron emission tomography with carbon 11 labeled Pittsburgh compound B. *Arch Neurol*. 2007;64:1140–1144.
13. Adamczuk K, De Weer AS, Nelissen N, et al. Polymorphism of brain derived neurotrophic factor influences beta amyloid load in cognitively intact apolipoprotein E epsilon4 carriers. *NeuroImage Clin*. 2013;2:512–520.
14. Bridging study of C11 PiB and F18 flutemetamol brain PET. ClinicalTrials.gov website. <http://clinicaltrials.gov/show/NCT01607476>. Accessed August 6, 2014.
15. Litchfield S, Nagy Z. New temperature modification makes the Bielschowsky silver stain reproducible. *Acta Neuropathol*. 2001;101:17–21.
16. Mirra SS, Heyman A, McKeel D, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology*. 1991;41:479–486.
17. Vemuri P, Whitwell JL, Kantarci K, et al. Antemortem MRI based STructural Abnormality iNDEX (STAND)-scores correlate with postmortem Braak neurofibrillary tangle stage. *Neuroimage*. 2008;42:559–567.
18. Lundqvist R, Lilja J, Thomas B, et al. Implementation and validation of an adaptive template registration method for [¹⁸F]flutemetamol imaging data. *J Nucl Med*. 2013;54:1472–1478.
19. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 2002;15:273–289.
20. The R Project for Statistical Computing website. <http://www.R-project.org/>. Accessed August 6, 2014.
21. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845.
23. Landau SM, Thomas BA, Thurfjell L, et al. Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. *Eur J Nucl Med Mol Imaging*. 2014;41:1398–1407.
24. Mosconi L, Tsui WH, Herholz K, et al. Multicenter standardized ¹⁸F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. *J Nucl Med*. 2008;49:390–398.